DEALING WITH HETEROGENEITY PROBLEMS AND CAUSAL EFFECT ESTIMATION IN ENTREPRENEURSHIP RESEARCH

ABSTRACT

This paper deals with causal effect estimation strategies in highly heterogeneous empirical settings such as entrepreneurship. We argue that the clearer used of modern tools developed to deal with the estimation of causal effects in combination with our analysis of different sources of heterogeneity in entrepreneurship can lead to entrepreneurship with higher internal validity. We specifically lend support from the counterfactual logic and modern research of estimation strategies for causal effect estimation.

1. EXECUTIVE SUMMARY

Entrepreneurship comes in many shapes and forms, driven by a broad variety of motivations and in a diversity of contexts. While this heterogeneity contributes to making entrepreneurship fascinating it also makes it very challenging for entrepreneurship researchers to arrive at strong and credible conclusions regarding causal relationships. Building on experiences from within and outside of entrepreneurship research this article provides an integrated discussion of strategies for dealing with the problem of heterogeneity with particular application to the entrepreneurship domain and the estimation of causal effects. Specifically, we deal with three problems: 1) *unobserved heterogeneity*, i.e., that unmeasured or unavailable variables may bias estimated relationships; 2) *causal heterogeneity,* i.e., that the structure, strength, direction or form of relationships may vary among sub-groups of the studied population, and 3) *uneven validity,* i.e., that the validity of chosen operationalizations may vary by sub-group or context. We discuss how these problems can be reduced at different stages of the research process, i.e., through *theory and theorizing;* in choosing a basic *design* for the study (including sampling)*;* at the *operationalization* stage, and through approaches chosen for *analysis,* respectively. We conclude each section with summarized advice that should help entrepreneurship researchers design more robust studies and arrive at more valid conclusions from extant data sets. Throughout, we illustrate with examples from entrepreneurship studies.

2. INTRODUCTION

This paper deals with causal effect estimation strategies in highly heterogeneous empirical settings such as entrepreneurship. Business ventures are started by individuals and teams with very different backgrounds and motivations, pursuing different objectives based on business ideas of very variable inherent quality in environments that also show tremendous variability. Certain aspects of this great variability or heterogeneity is an important, fundamental and theoretically interesting characteristic of the entrepreneurship phenomenon (Alvarez & Busenitz, 2001; Davidsson, 2004). After all, it is in great part their ability to deviate from norms in new and unexpected ways that makes new and growing ventures fascinating and – sometimes – financially successful. However, the great variability also makes it difficult for researchers to arrive at valid causal inference, and studies that try to 'reflect reality' by including all the multi-dimensional variance at once risk arriving at weak or confusing results. Similarly, studies seemingly addressing the same questions using different samples, operationalizations or analysis approaches may arrive at conflicting results. Over the years, this has led to frustration that "entrepreneurs seem to defy aggregation" (Low & MacMillan, 1988) and that we are "getting more pieces of the puzzle, but no picture is emerging" (Koppl & Minniti, 2003

In this article we use *heterogeneity* as an umbrella term for the simultaneous variability along three different dimensions that makes it challenging to adequately measure theoretical constructs and to correctly model and to estimate causal relationships. Numerous factors contribute to problematic heterogeneity. They are (1) unobserved heterogeneity, (2) causal heterogeneity and (3) uneven or differential validity.

While all scientific fields have to deal to some extent with heterogeneity problems there are several reasons why they are of particular significance for entrepreneurship research. First, the phenomenon itself may be more heterogeneous as it concerns emerging ventures, industries and populations. In more mature stages of development, market forces (Lawless & Tegarden, 1991), learning (Jovanovic, 1982) and institutional pressures (Henrekson, 2007) tend to limit the range of variation along at least some dimensions. Second, due to the cost and difficulty of obtaining primary data on such emerging phenomena researchers may turn to archival data that do not include all variables needed to avoid severe omitted variables problems (Shane, 2006). Third, within the multi-disciplinary field of entrepreneurship research, each theory or discipline emphasizes its specific set of variables and neglects others (Acs & Audretsch, 2003; Ireland & Webb, 2007). Partial absorption of what can be considered unified bodies of theory (and related empirical works) across a range of fields can lead to seriously misreading the theory and results. Many results may only be valid under certain theoretical assumptions and in specific empirical contexts. Hence, the construction of an integrated field of knowledge for entrepreneurship – important as it is – is associated with considerable risks[1].

This article provides an integrated account of the problem of heterogeneity related to causal effect estimation strategies as it presents itself through the entire research process and in an entrepreneurship context. We offer advice based on experiences from within and outside of entrepreneurship research. We also provide an entry point for entrepreneurship researchers to more specialized texts that cover in greater depth particular aspects of the heterogeneity problem from particular disciplinary perspectives but without integration or application to entrepreneurship (e.g., Hausman & Taylor, 1981; Rosenbaum, 2005; Shugan, 2006). We believe such a contribution is timely because evidence suggests that at the present time even the 'high end' of entrepreneurship scholarship struggles with heterogeneity problems and

[1] This is a paradox that entrepreneurship as field has to deal with (Ireland, Webb, & Coombs, 2005; Sorensen, & Stuart, 2008). As the field has grown it has moved towards fragmentation and research strongly grounded in theoretical perspectives has become more published (Ireland & Webb, 2007). This is good for the field and new and better knowledge is without doubt getting produced. However, it also makes it more difficult to integrate this knowledge because there is relative little academic premium to doing so compared to producing an empirical paper grounded in a single theory.

causal inference or the identification problem (Shane, 2006). In addition, we provide methods experts with a sense of the typical heterogeneity problems of entrepreneurship research, thereby facilitating their making additional contributions to increased sophistication of this field.

We organize our discussion as follows. In the next section we outline the various types of problems heterogeneity can cause and why this create a specific challenge to estimate causal effects. We then discuss remedies to heterogeneity problems through the use of theory and theorizing. We continue by addressing heterogeneity considerations at the design stage, including sampling strategies. After a discussion of how heterogeneity relates to operationalization we devote the second half of the paper to various remedies that can be applied when analyzing the data. Although we do not quantitatively review the occurrence of heterogeneity problems and specific remedies thereof in published entrepreneurship research we will throughout the manuscript make use of findings from previous method reviews. We will also provide illustrations of how selected entrepreneurship studies have successfully dealt with the heterogeneity issues we raise.

3. HETEROGENIETY AND CAUSAL EFFECT ESTIMATION

Most social research studies, whether quantitative or qualitative, deal explicitly or at least implicitly with causal relationships (King, Keohane, & Verba, 1994). Entrepreneurship is no exception, as illustrated, for example, by Shane and Venkataraman's (2000: 218) three fundamental questions for entrepreneurship research. Entrepreneurship researchers takes an interest in understanding what personal characteristics make individuals engage, persist or succeed in business start-up activities (Davidsson & Honig, 2003). Alternatively, they seek explanations for differential levels of innovativeness in the characteristics of the firm itself (Cliff, Devereaux-Jennings, & Greenwood, 2006) as well as in the conditions of its regional environment (Maillat, 1998). In other cases still they may want to understand how national institutional conditions influence levels and contents of entrepreneurial activity across countries (Henrekson, 2007).

These examples concern how one or more circumstances or factors ('explanatory variables') cause one or more outcomes ('dependent variables'). In order to illustrate the type of heterogeneity problems encountered in such research we can start with the simple model displayed in Figure 1. In this figure, X denotes theoretical constructs whose influence is to be assessed on the dependent variable(s), i.e., theoretical construct Y. The operationalizations of these constructs are denoted by *X'* and *Y*', respectively. The solid arrow from X to Y represents the true relationship between the theoretical entities whereas the dotted arrow between *X'* and *Y'* represents the estimates obtained in the research (cf. Bacharach, 1989). We may think of X as the variables whose causal influence on Y we have a theoretical interest in. However, in order to avoid biasing influence of heterogeneity we may also want to include other variables in *X'*. Z represents variables that are not of theoretical interest as disregarding them may lead to a biased picture of X → Y relationships. Some Z variables may be measurable/available and thus possible to include in *X'* whereas other Z variables may be genuinely unobservable.

++++++++++++++++++++++++++

Insert Figure 1 about here!

++++++++++++++++++++++++++

In the situation illustrated by Figure 1 an array of problems may lead to weaker or less correct explanation of Y than expected. Some of these problems are a) low explained variance in *Y'* because of the exclusion of Z; b) low explained variance in *Y'* because (uniformly across the studied population) causation is probabilistic (i.e., imperfectly regular; cf. Yang, 2006); c) low explained variance in *Y'* and under estimated influence of X because of uniformly low validity and/or reliability (large measurement error) of *X'* or *Y'* or both, or d) unreliable estimates of the influence of individual X variables due to high inter-correlations among X. However, neither of these problems are instances of systematic bias due to heterogeneity. In the following, we limit the discussion to problems of the latter kind, namely:

#1. *The problem of unobserved heterogeneity* (e.g., Shugan, 2006). Also discussed under labels such as *omitted variable bias* (Lee, 1982) or *confounding variables* (Kish, 1987), this is the central heterogeneity problem that if Z has substantial correlations with both *X'* and *Y'*, excluding Z will lead to serious bias in the coefficients. The problem is aggravated if the omitted Z variables are causally related with X as the model is then also structurally misspecified (see. #2).

#2. *The problem of causal heterogeneity* (e.g., Western, 1998). Related to the notion of *boundary conditions* (Bacharach, 1989) and with mediation and moderation modeling as solutions in special cases (Baron & Kenny, 1986), this concept denotes the more general problem that the effect of X on Y may not be uniform across elements or subgroups of the studied population. For example, the effect of one X variable may be different in different subgroups, contingent on the value of another continuous X variable, or affect Y indirectly via another X variable.

#3. *The problem of uneven validity. uneven validity*Different aspects of this problem are highlighted particularly in cross-cultural research under labels such as *construct equivalence, instrument equivalence, measurement equivalence* and *measurement invariance* (e.g., Byrne & Watkins, 2003; Schaffer & Riordan, 2003; Singh, 1995). If the X ←→ *X'* and/or Y ←→ *Y'* correspondence – i.e., validity – is uneven it means the chosen constructs and/or operationalizations are not equally suitable to all subgroups of the population. As a result *X'* → *Y'* relationships will be misestimated. That is, this method artifact may be misinterpreted as substantive differences in the nature and strength of relationships.

The omnipresence of these heterogeneity problems may make it impossible to conduct a flawless study of entrepreneurship. However, a researcher who does not try hard to approach that ideal is unlikely to come up with strong and valid findings about the phenomenon. In the remainder of this manuscript we will discuss how these sources of heterogeneity can be dealt with through decisions related to theory, design, operationalization, and analysis.

4. DEALING WITH HETEROGENIET THROUGH THEORY AND THEORIZING

Theory is usually defined as "constructs linked together by propositions that have an underlying, coherent logic and related assumptions" (Davis, Eisenhardt, & Bingham, 2007: 481; cf. Buchanan, 1989: 496, 498) or some variation on that theme. In this section we will show how some well known frameworks regarding theory can be applied in order to address heterogeneity problems in entrepreneurship research. First, we approach the level of maturity in our field as it has an effect on not only how models are constructed and understood, but also how we can handle heterogeneity problems. Second, we explain the functioning of the increasingly popular counterfactual logic. Third, we discuss the specific causal effect estimation strategies derived from the counterfactual logic and that have been developed the last ten years.

4.1 Mature and Nascent Theories

When the constructs, their links and the underlying logic are well specified – the mature theory situation according to Edmondson and McManus (2007; cf. Zahra, 2007) – it follows directly from theory what variables need to be included and how their relationships should be modeled. If the theory is well established its boundary conditions should also be well established, meaning that the theory will also indicate what contexts or samples should be studied. In the extreme case all heterogeneity problems and their solutions can be derived directly from the theory. At the other extreme Edmondson and McManus (2007) put the ‘nascent’ theory situation. When the theoretical knowledge of the phenomenon is rudimentary or non-existent, increased familiarization with the phenomenon through qualitative work is likely to be needed in order to even begin building an understanding of what problems of unobserved and causal heterogeneity might exist.

We would argue that in most entrepreneurship research situations the starting point is somewhere between these two extremes (cf. Zahra, 2007). The challenge is then to systematically tease out what heterogeneity exists and determine which aspects of it are methodologically troublesome and theoretically interesting, respectively. It has been argued that “Ideal theory tests should only include the variables in the theory” (Shugan, 2006: 203). Hence, variance deriving from sources outside the theoretical domain needs to be excluded or controlled for because otherwise it leads to biased estimates of causal relationships and hampers theory development. We therefore need tools that allow us to generate theory about causal effects and tools that allows to transform theory is to effective research designs where twe can make the best use of our current knowledge knowing that it is only partial.

4.2 Identifying Sources and Specific Effects of Heterogeneity

Counterfactual argumentation or the potential outcome model is an increasingly popular way to generate theory and research designs that are better able to establish causality (Pearl, 2000). Counterfactual argumentation is a logic of inference that plays a central role in establishing causality in various fields such as political science (Fearon, 1991), history (Lebow, 2000), medicine (Höfler, 2005) and economics (Heckman, 2000).

++++++++++++++++++++++++

Insert Figure 2 about here

+++++++++++++++++++++++++

As a starting point we can assume a theory making a proposition about an X→Y relationship where we would like to estimate the size of the causal effect. A counterfactual argument is based around the question if a specific event E would had happened if the cause C would have not been present. That is, "if it had been the case that c (or not C), it would have been the case that E (or not E)": Counterfactual make claims about events that did not occur. Such propositions play an essential and elementary role in the efforts of the social sciences to assess their hypotheses about the causes of the phenomena they study. The counterfactual argumentation becomes especially important when we no longer can rely on experimental design (Pearl, 2000).

Defining causality when the causes are interrelated, partially unknown, perhaps immeasurable, and data is mostly based on observation is less straightforward and becomes a major achievement. The use of counterfactuals in theory and hypothesis testing is a way to mitigate this problem. The use of counterfactual arguments is important because they allow us to better handle the problem that many models can explain the same data, and that is the formal arguments in the model that decide causality. Differently stated, in social sciences we can observe the same causal patterns, but we can imagine different interpretations to this causal patterns. The logic of inference that is counterfactuals allows us to more clearly state different theoretical explanations and under what conditions they are likely (and not) to explain what we observe. It is through its use of logic that we can say what variables need to be present to fully examine the causality argument or treatment effect and what variables can be excluded to create a parsimonious models. The use of counterfactuals makes a clear link between the nature of theory, model and empirical testing and they are mutually dependent.

The main advantage with the counterfactual arguments is that the researcher in explaining why some particular event E occurred cannot help to explain why E occurred rather than another possible outcome or outcomes. The researcher need to be clear on what the specific cause is, why it should affect the particular outcome and what would happen in the absence of the cause being present. The later is the counterfactual argument. The counterfactual argument is only as compelling as the logic and "evidence" offered by the researcher to verify the links between the hypothesized antecedent and its expected consequences (Lebow, 2000). The counterfactual not only develops a logic of argumentation in the theory, it also leads to imagine a research strategy that provides the "empirical" confirmation for a causal hypothesis.

Let us assume that we want to test the following hypothesis: "C is a cause of event E". There are two complementary approaches to test this empirically. The first approach is to imagine that C has been absent and ask whether E could have (or might have) occurred in that counterfactual case. This leads to four different potential outcomes depicted in Figure 2 that needs to be examined in our models. For example, we can imagine a theory suggesting that unemployment increase the likelihood of starting a businessbecause those faced with unemployment have reason to seek alternative ways to provide for themselves. Quadrants I and III denote the cases that accord with this explanation– when X (employment status) changes, Y (probability of creating a start-up) changes as well (Quadrant I). When there is no change in X, no change in Y is observed (Quadrant III) . Quadrants II and IV constitute the 'counterfactual' cases. In quadrant II the question is under what conditions the proposed relationship might not hold, i.e., why does getting unemployed not lead to an increased probability engagement in self-employment. Quadrant IV depicts the final outcome: why do business start-ups while not being unemployed? Hence, we constructs a space of possible outcome with related probabilities. Theory is there to explain under which circumstances and conditions some people are more likely to end up in one quadrant than another. Basically what are the sources of heteroegeniety we need to deal with to test our hypothesis that unemployment leads to the creation of new start-ups.

Whetten's (1989) analysis of the components of theory is one possible guide to identifying sources of heterogeneity problems. Referring back to our definition of theory, in Whetten's exposition constructs constitute the *What?* of theory, whereas propositions concern the *How?* The coherent logic he associates with *Why?* whereas the *Who?*, *When?* and *Where?* concern the assumptions and boundaries of the theory's applicability. Although Whetten (1989) does not think of the *Who?*, *When?* and *Where?* as the likeliest candidates for strong theoretical contributions we would argue these questions provide a good starting point for to determine what heterogeneity might exist and determining what aspects of it to try to include and exclude, respectively, in the design. We hold this view because boundary conditions of theory are often not satisfactorily specified (Davis et al., 2007).

The questions *Who?*, *When?* or *Where?* are largely questions about the role of context, i.e., under what contextual conditions theoretical relationships are likely to hold or vary. They therefore not only inform about sources of heterogeneity, but also on the possible effects of that same heterogeneity. Johns (2006) provides a very useful discussion of what context does to empirical relationships. In short, context a) restricts range; b) affects base rates; c) changes causal direction; d) reverses signs; e) prompts curvilinear effects, and f) tips precarious relationships. This is largely about causal heterogeneity across contexts. Johns (2006) also adds a category g) 'threatens validity', which may seem superfluous as all of the above also threaten various aspects of validity. However, when the meaning or conceptual relevance of a construct itself – a *What?* of the theory – varies, we have a case of lacking 'construct equivalence' (Singh, 1995); i.e., uneven validity on the theoretical level. Contextual consequences in terms of construct inequivalence as well as causal heterogeneity highlight the need to carefully consider context in theorizing, and therefore also in design and analysis strategy.

Applying Whetten's (1989) and Johns' (2006) notions to our hypothetical case about unemployment and business start-ups several quadrant II possibilities present themselves. In a society or social stratum where most individuals are affluent by birth the construct 'unemployed' would not be equivalent with the same notion in mainstream societies. Where the institutional framework includes generous unemployment benefits and/or high bureaucratic barriers to firm formation, the relationship could be weak or non-existent. For people close to retirement age the response to lay-offs may more rarely be to set up their own business. Importantly, if the theoretically focused variable is correlated with another variable that has a negative effect on Y, the positive effect of X will not necessarily appear in empirical estimation. This would be the case here if the economic conditions that increase unemployment and therefore increase necessity-based entrepreneurship at the same time reduce opportunity-based entrepreneurship via decreased market demand (cf. Wennekers, Stel, Thurik, & Reynolds, 2005; Hamilton, 1989). These added insights into the possible nature of the relationship have consequences for sampling, variable inclusion, and modeling as further discussed in later sections. It should also be clear from the example that the counterfactual mental gymnastics has potential for better defining the boundary conditions of the theory or for enriching it with new contingencies., i.e., the exercise has repercussions on more fundamental aspects of the theory because, in Whetten's words "theorists need to understand why this anomaly exists, so that they can revise the *How* and the *What* of the model to accommodate this new information" (Whetten, 1989: 493). We would argue that the repercussions extend to the *Why?* question as well.[2]

Obviously, like in most social research we are dealing with a phenomenon that has many possible causes. This is the root of the problem of unobserved heterogeneity. These other causes pose problems when they are correlated with the cause we take a theoretical interest in. Therefore, the essence of systematic search for other causes is to find those correlated variables that either need to be included in the empirical design or made uncorrelated with X (or, more correctly, with *X'*) via constant-holding, randomization, or matching. If they are genuinely non-measurable, we may still want to assess the potential bias they cause through indirect means. Again, this will be further discussed in sections to follow.

4.3 Strategies for Causal Effect Estimation

It should now be clear that we are more interested in identification than in statistical inference. We have made two statements so far. First, entrepreneurship theories are in general not very mature. This means that we have limited knowledge of the different causal mechanisms that we are interested of. Second, we have stated that counterfactual thinking is a useful tool to think about how causality can be inferred from observational data. Our third argument is derived from the second as we here introduce the seminal work of Judea Pearl (2000), who developed a set of rules for representing causal relationships with graph theory. We do not aim to present a general introduction to his work. We only aim at introducing the most basic elements of his work in order to point towards the enormous strength of his work to help us develop better strategies for causal effect strategies.

Pearl (2000) use graph theory instead a mathematical notations because it provides a better understanding of the causal relationship to be studied, as well as it provides some

[2] For example, it is conceivable that an idealistic environmentalist currently in gainful employment and a profit-maximizing entrepreneur currently active in another industry both respond to a given institutional change by shifting from what they did previously to trying to launch similar, environment-saving new ventures.

important tools to decide what variables to include in the model and why. The idea is that we have a universe of variables that affect X and Y that be represented graphically. Given this graphic representation of the causal structure which variables must we observe and then use in the data analysis to estimate the size of the causal effect of X on Y?

The use of graph theory leads to two important advantages that allows us to better answer to the above question. First, his framework is completely nonparametric. It is therefore not necessary to specify the nature of the causal relationship between X and Y (linear, quadratic or something else). X just causes Y. This eliminates the Achille's heel of traditional path analysis which needs to make this kind of assumption. Causal models therefore become easier to handle. Second, Pearl shows that exist three basic strategies to identify a causal effect. They are: conditioning on variables that block all back doors (e.g., stratification, matching, weighting, regression), conditioning on variables that allow for estimation by a mechanism (mediation analysis), and estimating a causal effect by an instrumental variable that is an exogenous cause to the cause.

The first strategy is to eliminate sources of unwanted heterogeneity by invoking Pearl's back door criterion using available theory and measure to control for these sources (causal heterogeneity and differential validity). That is, all back-door paths from the causal variable to the outcome variable are controlled for. This strategy is called simple conditioning. The second strategy of mediation is to establish an isolated and exhaustive mechanism that relates the causal variable (X) to the outcome variable (Y) and then calculate the causal effect as it propagates through the mechanism. The third strategy is to use an exogenous instrumental variable to isolate covariation in the causal and outcome variables (Morgan & Winship (2007).An important decision criteria in how to model and test the causal relationship are the sources of heterogeneity . The most well-known and used strategies is simple conditioning. However, the point is that the use of graph theory combined with current knowledge allows the researchers to better estimate what variables to exclude and include and why, and consequently which are the best design, measurement and analytical approaches to follow.

In this section we have discussed how theory and theorizing can be used to deal with heterogeneity problems; in particular the problems of unobserved heterogeneity and causal heterogeneity. In Table 1 we summarize our advice.

++++++++++++++++++++++++++++

Insert Table 1 about here!

++++++++++++++++++++++++++++

5 DEALING WITH HETEROGENEITY IN DESIGN

It should be clear from the many sources and forms of heterogeneity discussed so far that the 'obvious' solution to this heterogeneity problem – to include all influential variables, measure them correctly and find the right model for all of their interrelationships – simply is not achievable. Neither may it be desirable, in the interest of parsimony and simplicity. A reason to prefer simpler theories is that such theories are more constraining and thus more falsifiable. A parsimonious model is not over fitted, which increases its credibility if it accords with the data (Pearl, 2000; Popper, 1959). Hence, design aims at creating a situation where the influence of one particular variable or a limited set of variables can be studied without the potentially disturbing influence of other variables. We discuss three strategies for approaching that ideal: 1) applying experimentation and other controlled approaches using 'artificial' data; 2) exploiting experiment-like real world situations, and 3) reducing heterogeneity through sampling from one or more sub-populations characterized by relatively high internal homogeneity. These three research designs are example on how to conditioning on the variables that blocks the back doors of the casual relation to be estimated. We here assume a good theoretical knowledge and the availability of good measures. When unobserved heterogeneity is a problem, then the use of longitudinal designs allow for an important safeguard. Longitudinal designs are discussed in some details in the section on related analysis techniques.

5.1 Experimentation and Other Laboratory Research Methods

Reducing heterogeneity lies at the heart of experimental research. Through manipulation of the focal explanatory variable(s) ('treatments') and randomization or constant-holding of other influences the experimenter can eliminate the problem of unobserved heterogeneity. In practice, experimentation is likely to reduce problems of causal heterogeneity and ununeven validity as well because of the reduced complexity of the research setting. Yet, the use of experiments has been limited in entrepreneurship research. Chandler and Lyon, (2001) and Bouckenooghe et al. (2007) both report 4 percent of the studies reviewed used an experimental approach with perhaps as little as 1 percent being based on 'laboratory' experiments.

With some creativity a range of entrepreneurship issues can be addressed through experiments and other methods using laboratory control (see Baron, 2006; Baron & Ward, 2004; Schade, 2005). For example, Dean Shepherd has championed empirical work using conjoint analysis and other experimental approaches to address otherwise hard-to-study issues such as opportunity recognition and effects of emotions (Brundin, Patzelt, & Shepherd, 2008; Gregoire, Barr, & Shepherd, forthcoming; Shepherd & DeTienne, 2005; Shepherd & Zacharakis, 1997). Similarly, Gustafsson (2006) compared expert and novice entrepreneurs through experimentation, largely supporting Hammond's (1987) theoretical proposition that experts have the ability to adapt their decision-making style to the level of uncertainty associated with a particular venture, while novices tend to apply the same approach regardless of task characteristics. These experimental studies have in common that they produce relatively clear answers that are likely to be replicable. Further, Sarasvathy (2001, 2008) developed Effectuation Theory through laboratory work – non-experimental but retaining many of the heterogeneity reducing characteristics of experimental work as she had expert entrepreneurs think aloud about the same opportunity rather than studying each pursuing their own, idiosyncratic venture.

Simulations offer another type of controlled context for developing and testing theoretical ideas. According to Davis et al., (2007: 483) “simulation is especially useful for theory development when the focal phenomena involve multiple and interacting processes, time delays, or other nonlinear effects such as feedback loops and thresholds” – a characterization we would argue fits well with the reality of entrepreneurship. The studies by March (1991) and Nelson and Winter (1982) are famous examples of how this approach has been applied to topics that are entrepreneurship-related. A ‘narrower’ example close to the core of entrepreneurship is provided by Fiet, Piskounov and Patel (2005). We would encourage increased use of experimentation and simulation in entrepreneurship research, at least as steps towards ascertaining internal validity in a ‘full cycle’ approach to building evidence. In such an approach, the researcher aims at first identifying relevant problems in the field and then reducing the characteristics of the problem so that theoretical relationships can be tested in a ‘laboratory’ setting. Once affirmative evidence has been achieved under such controlled conditions verification in more real-life like setting can be sought (Cialdini, 1980; Chatman & Flynn, 2005). This approach can help avoid lack of progress in research on new questions arising from conflicting results which in turn are due to overwhelming heterogeneity problems.

5.2 Exploiting Experiment-Like Situations

For practical or ethical reasons most entrepreneurship research problems cannot be studied experimentally. For example, we cannot get governments to systematically vary the institutional framework across regions according to an experimental design, make people and businesses stay in the regions they have been randomly assigned to, and expect them to behave as if were it a uniform, permanent institutional change rather than an experiment. That is, for the most part entrepreneurship researchers have to rely on observational studies. These differ from experiments in that the investigator cannot control the assignment of the treatments to subjects (Rosenbaum, 2002). One partial remedy – in particular to unobserved heterogeneity problems – is then to work on a well defined population that exogenously receives the same treatment. Under such circumstances members of the population cannot choose whether or not they receive the treatment, but they can choose how to respond to it. Hence, there is no problem of self-selection (Heckman, 1979; 2000).

One elegant example that emphasizes how a group of dissimilar entities react to the same 'treatment' is Shane's (2000) study of all ventures (and their founders) associated with one and the same basic technological innovation. By keeping the basic innovation constant and including the entire 'population' (all of whom came from the same university) the risk of unmeasured heterogeneity distorting the results could be reduced. The study provides compelling evidence that prior knowledge is very important in determining what specific business opportunities entrepreneurs will discover and/or succeed with under otherwise equal conditions.

Tragic as they are, natural disasters like Hurricane Katrina and the Boxing Day Tsunami also provide entrepreneurship research opportunities (see Dickson & Kangaraarachchi, 2006; Runyan, 2006). For example, comparisons of entrepreneurs (business owners or founders) with others (managers or general population) typically confound factors that make people *engage* in entrepreneurial endeavors with those factors that make them *persist* and *succeed,* respectively, at such tasks (Davidsson, 2004: 70). Post-disaster situations present a cleaner context for addressing the specific issue of entrepreneurial persistence. Galbraith and Stiles (2006) present some results on that particular issue.

Natural situations where cases vary on one variable that can be assumed to be uncorrelated to other variables, similarly to randomization in experiments, are a variation on this theme. An example here is research on the impact of windfall gains on entry into self-employment. This research is interested in how liquidity constraints hamper entry into self-employment. Lottery wins are randomly distributed among participants in the lottery, so there is no problem of self selection and the researchers can argue that the treatment effect is exogenous to the model. Lindh and Ohlson (1996) and Taylor (2001) both found that winning at the lottery increases the probability of entering into self-employment. The significance of this finding should be understood in the context of much other – and less controlled – observational research on entrepreneurial propensity finding a surprising absence of effects of financial variables (Davidsson, 2006; Kim, Aldrich, & Keister, 2006).

5.3 Narrow Sampling

While natural experiments can be creatively and opportunistically capitalized on for shedding light on some issues, they are unlikely to serve the purpose of informing the particular issues the researcher already had an interest in. When experiments and experiment-like designs are not possible the researcher can reduce bias due to unobserved heterogeneity by using a narrow sample like a single industry, a narrow age cohort or size band, or a combination of these. The main reason for this is that a range of variables Z that would influence Y in a more broadly based sample will be constants or near constants in the more restricted context, thus not contributing to variance in Y. In addition, a less complex empirical setting facilitates familiarization with the studied phenomenon, which arguably reduces the risk that relationships are misspecified for parts of the studies population, i.e., the problem of causal heterogeneity, as well. As we will discuss later, narrow sampling also helps reduce the problem of uneven validity.

Sampling narrowly follows the logic that if the theory is robust, it should make correct predictions for this narrow, homogenous sample. Applicability of the results beyond the studied context would rely on idenfication rather than on statistical inference future replications in other contexts. Recent examples show that theory-driven research on narrow samples can lead to strong results regarding relationships that have appeared weak or inconsistent in previous research. Exemplary in that regard is Baum and Locke's (2004) study of how individual psychological traits influence the growth of young firms. They use a sample not from the entire small business population but a much narrower category: North American architectural woodwork firms[3]. While excellent research craftsmanship regarding other design features also has part in the relative success of this study the decision to design away potentially blurring heterogeneity by focusing on a narrow industry was no doubt an important contributor.

In fact, entrepreneurship research that achieves publication in top tier disciplinary or mainstream management outlets is often of this kind: the researchers have identified a homogenous sample that allows them to test their ideas in a more controlled setting. Other examples are Eisenhardt and Schoonhoven (1990) semiconductors), Stuart et al. (1999; biotech) and Usher and Evans (1996; gas stations in Calgary). In neither case was the research driven by an interest in these industries or particular geographical contexts. The contexts were chosen because they presented one relevant arena for testing theoretical ideas, and to do so without blurring issues by including variation along too many dimensions at once.

The gain of internal validity comes at the possible cost of generalizability (Rosenbaum, 2005). Arguably, the severity of this concern varies across studies. Regarding Baum and

[3] This is a context which is not only narrowly defined but also one with which the first author of the study was highly familiar; see www.highmark.com/hmk2/about/boardofficers/baum.shtml.

Locke's (2004) results it is difficult to see why this type of variables would have completely different effects in other industries, countries or periods. However, is Cliff et al.'s (2006) finding that insiders at the periphery of the industry rather than either its core firms or industry outsiders are the most innovative – based on a study of organizational innovation in law firms founded in Greater Vancouver – representative for other forms of innovation? And are law firms good representatives for the theoretical category 'firms' or 'young firms'? The authors themselves see them as representative of highly institutionalized, mature industries and found their selection suitable also because it was easy to identify 'core' and 'periphery' of the industry. The relative importance of outsiders vs. insiders may well vary with industry maturity as well as with type of innovation, as the authors admit when discussing the limitations of their research.

5.4 Stratified Sampling and Sample Equivalence

We have already noted that while narrow empirical contexts are good for theory testing, they do not allow for statistical generalizations to broader populations. Statistical generalization requires random sampling (or, more precisely, probability sampling) from the population to which inference is to be made (Oakes, 1986). However, relying on a random sample of 'all firms', 'all entrepreneurs' or 'all start-ups' may not be a good strategy for any purpose. A simple random sample may be totally dominated by, e.g., tiny, imitative businesses in mature industries. The mere fact that they are more numerous in a given country at a given time does not necessarily make them more important from a theoretical (or economic) point of view. Further, while increasing the size of a simple random sample from a broadly based population reduces random sampling error it does nothing to reduce the three heterogeneity problems we are discussing.

We argue that the issue is not about a choice between random vs. purposive sampling but about attaining the ideal of *purposive random sampling*. That is, the more important issue may be to ensure satisfactory representation of different types of firm or context, while the idiosyncrasies of non-selected categories are excluded by design[4]. If sampling within each theoretically relevant stratum can be done randomly, statistical inference theory remains applicable. This is a major advantage (Oakes, 1986).

Notions from cross-cultural research (Schaffer & Riordan, 2003) are useful for making decisions about stratified samples. Such research distinguishes between 'emic' and 'etic' approaches. In the former case the cultural concepts and theoretical predictions are essentially derived from one culture and their occurrence may (or may not) then be compared with other cultures. The 'etic' approach, which dominates cross-cultural research, assumes that shared frames of reference exist and requires in-depth knowledge of each compared culture in order to avoid bias. Schaffer and Riordan (2003) propose a combined emic-etic or a derived etic approach as best practice. The equivalent for entrepreneurship research would be to take great care in theory development and pilot testing so that the design is not implicitly based on conditions in a particular industry or type of firm when intended to be valid across sampled strata (cf. Kleinknecht, Van Montfort, & Brouwer, 2002 on 'manufacturing bias'). Another useful notion from cross-cultural research is that of 'sample equivalence' (Coviello & Jones, 2004; Schaffer & Riordan, 2003). For example, in the *Global Entrepreneurship Monitor* research it can for some purposes be called in question whether the different country samples of 'nascent entrepreneurs' represent the same phenomenon when some countries are dominated by 'necessity-based entrepreneurship' while in others 'opportunity-based entrepreneurship' (Wennekers et al., 2005) predominates.

Apart from the experimentation- and sampling issues discussed above there are additional design issues relating to longitudinal data collection and inclusion of control variables. Table 2 provides a summary of design-related measures to reducing heterogeneity

[4] In this regard, results from recent reviews suggesting that some 33-44 percent of entrepreneurship research published in top tier journals rely on random sampling (Bouckenooghe et al., 2007; Ireland et al., 2005) are not directly indicative of whether this minority represent better research practice.

problems, including also these latter issues. However, we further discuss longitudinal design and analysis to a later section on strategies for dealing with heterogeneity in the analysis. For reason of space and that the use of control variables are relatively well-understood we do not go deeper on this subject here.

++++++++++++++++++++++

Insert Table 2 about here!

++++++++++++++++++++++

6 DEALING WITH HETEROGENIETY IN OPERATIONALIZATION

6.1 Uneven Validity

The heterogeneity of the entrepreneurship phenomenon also leads to the problem that a chosen operationalization may not be uniformly valid across the studied population. Consider the *Entrepreneurial Orientation* (EO) scale (Lumpkin & Dess, 1996), which is no doubt the most frequently recurring measurement instrument specific to entrepreneurship (and which has yielded meaningful results in many studies; cf. Rauch et al., 2006). One of the items of the innovation sub-scale is “How many new kinds of products or services has your company introduced over the past 5 years?” How would ‘equally innovative’ manufacturers and service firms answer this question? As a way to partly overcome the problem of general applicability the high end of the response scale employs a weak quantification: ‘a lot of new products/services.’ This makes it highly subjective in addition to still being sensitive to industry and firm size, and as regards retailing firms it is questionable whether the item is meaningful at all.

This problem of uneven validity is well known in cross-cultural research (Byrne & Watkins, 2003; Coviello & Jones, 2004; Schaffer & Riordan, 2003; Singh, 1995) and this is, again, a suitable place to turn for inspiration. On the operational level, Shaffer et al. (2003)

further subdivide this problem into *conceptual equivalence* (that survey items elicit the same conceptual frames of reference) and *scaling equivalence* (that respondent categories perceive and interpret rating-scale intervals in the same manner). As best practice they propose the use of covariance structure analysis and item response theory. Byrne and Watkins (2003) as well as Singh (1995) provide detailed procedures for assessing and remedying various aspects of uneven validity.

Although the problem of assessing 'the same' issues across different industries and types of firm arguably corresponds to the problem of uneven validityuneven validity encountered in cross-cultural (cf. Byrne & Watkins, 2003: 155) research it is rarely explicitly discussed or systematically dealt with in entrepreneurship research. For example, we have found no published study specifically testing for uneven validity of the EO scale other than the conventional country comparisons (Kreiser, Marino, & Weaver, 2002). However, for a similar, firm level measure based on the notion of *Entrepreneurial Management* (Stevenson & Jarillo, 1990) it was demonstrated that the factor structure held up in each studied sub-population by industry, firm size and governance structure (Brown, Davidsson, & Wiklund, 2001).

A particular problem is when the dependent variable is not equally applicable across the studied population. For example, Venkataraman (1997) argued that specific outcome measures might not be equally relevant to all firms and therefore not meaningfully comparable. That is, specific outcome measures may not have conceptual equivalence if interpreted as indicators of 'performance' or 'success'. An early review by Murphy, Trailer and Hill (1996) indicated this is a widespread problem in entrepreneurship research. Essentially, their analysis suggests that a) a single dimension may not be meaningfully comparable and b) a multiple indicator index may not be a good solution, either. They therefore suggest that "future studies [of performance in an entrepreneurship context] should, where possible examine multiple dimensions and use multiple measures of those dimensions" (Murphy et al., 1996: 22). A recent review by Bouckenooghe et al. (2007) indicates this call has so far not been sufficiently heeded by entrepreneurship researchers. According to their compilation a majority of studies that had an identifiable DV applied only one measure and only 10 percent of all studies used four or more dependent variables.

6.2 Narrow Sampling, Uneven Validity, and Replication

Focusing on a narrow sample also brings operationalization advantages. First, since many variables will become constants more questionnaire space or interview time can, *ceteris paribus,* be allotted to the central constructs of the study. Second, since equivalence issues are designed away, highly customized and therefore more precise operationalizations can be used. The previously mentioned study by Cliff et al. (2006) illustrates this. Because they only studied law firms they could create an overall index of innovativeness by assessing how 15 specific practices deviated from the industry norm. Arguably, this resulted in a more valid measure than a generic alternative designed to serve equally well for assessing innovativeness among, e.g., manufacturers and retailers as well. Similarly, when Harrison et al. (2004) investigated financial bootstrapping behavior using a sample from the software industry, they included items like 'commercializing public domain software'; 'porting fees to transfer software from one platform to another' and 'using public domain development tools' alongside more general indicators.

In both cases, the improved internal validity comes at the cost of developing a measure that cannot be applied elsewhere. This raises the question of how to allow for replication. We agree with Hubbard, Vetter and Little (1998) that replication is both critically important and sadly under utilized in management research, including entrepreneurship studies. However, we would also argue that contexts for entrepreneurial phenomena are so heterogeneous that in many cases an exact replication on the level of operationalization is not feasible. Replication on the abstracted level of theory may be a more sensible route. That is, comparison of the same X → Y relationship in Context A and Context B conducted as two separate studies, with operationalizations and controls adapted to each context may be preferable to trying to investigate the relationship with the exact same operationalizations and controls in both contexts.

Our advice concerning dealing with heterogeneity in operationalization is summarized in Table 3.

++++++++++++++++++

Insert Table 3 about here!

++++++++++++++++++

7. DEALING WITH HETEROGENEITY IN ANALYSIS

Whether or not heterogeneity has been reduced by design, a number of approaches can be used to deal with it in analyses which are related to Pearl's three basic strategies to deal with the estimation of causal effects. . This is what we turn to next. We start with approaches to causal heterogeneity where simple conditioning is possible which can be used with both cross-sectional and longitudinal data (moderator variables and matching techniques). Second, we investigate techniques that are available to researchers when simple conditioning is no longer effective. We here explore mediation and instrumental variables which represents the two remaining strategies proposed by Pearl. Finally, we then move on to techniques that offer some of the strongest safeguards against unobserved heterogeneity (sensitivity analysis and longitudinal analyses).

7.1 Analysis Techniques for Dealing with Causal Heterogeneity

We here review two different techniques that can be derived from the basic technique using simple conditioning which addresses foremost causal heterogeneity and to a certain degree uneven validity. The two techniques are moderator or interaction analysis and matching techniques.

*Models with moderator variables***.** There are different ways to model an expectation of differential effects of one explanatory variable depending on the value of another explanatory variable (the so called 'third variable'). The idea is that this third variable when combined with our primary explanatory variable X will change the relationship altogether between X and Y (Baron & Kenny, 1986).

A moderator or interaction effect exists when the effect of the independent variable on the dependent variables differs depending on the value of a third, moderating variable (Aguinis, 2004; Jaccard & Turrisi, 2003; ORM, 2002). Inclusion and correct modeling of moderators is a key strategy to solving the problem of causal heterogeneity. Ketchen, Boyd and Bergh (2008) show that its use has increased dramatically in management research. The same appears to be true for entrepreneurship research. Almost every issue of the 2006-8 volumes of the leading niche journals (*Journal of Business Venturing* and *Entrepreneurship, Theory & Practice*) include one or more articles applying some form of analysis of interaction effects. For example, Walter, Auer and Ritter (2006) demonstrate that the positive effect of Entrepreneurial Orientation on performance is in part contingent on network capabilities. Zahra and Hayton (2008) similarly show that a positive effect of international venturing on firm performance is contingent on the firm's absorptive capacity.

There is much discussion about how to best test for and interpret moderator effects. Aguinis (2002) cautions that popular methods for estimating interaction effects are fallible and Brambor et al. (2006) found in a review of three major journals in political science that only 10% of the 156 reviewed papers made correct interpretations. Our overview of the two above-mentioned journals indicates that entrepreneurship may not be any better off. Hence, entrepreneurship researchers should consult cutting edge expertise before interpreting and publishing such findings. Following the example of previously published entrepreneurship research does not guarantee correct application or interpretation.

Matching techniques and propensity score methodology. Post-matching techniques rest on the basic idea that an almost experimental design with a control group and a treatment group can be created while still using observational data if we can eliminate differences in individuals prior to being exposed to the treatment. This alleviates the problem of self-selection and it is based on the conditioning on the back door variables. Through matching the estimation of the causal effect of the said treatment is improved. A popular recent technique is the propensity scores method (Rosenbaum and Rubin, 1983). It is broadly applied in areas such as medicine and economics (Heckman & Navarro-Lozano, 2004).

We describe the technique with an entrepreneurship example. Nopo and Valenzuela's (2007) study whether or not individuals who switch to self-employment receive an increase in income. The switch here represents the treatment. The analysis is complicated by the fact that there is a range of observable and unobservable variables that determine whether or not a person engages in self-employment, and those might also have an effect on the ability to generate income. To isolate the effect of switching and control for these potentially disturbing variables, the authors use propensity scoring. They create matches to all persons entering into self-employment. These matches build on all comparable characteristics, but this is converted into a single variable, the propensity score. This makes it possible to estimate the average treatment effect for entering into self-employment. They find positive effects for both men and women on income, the latter effect being somewhat stronger.

Another example is Wang's (2008) study of how a property right reform in China in the mid-1990s affected entry into self-employment. The treatment is the reform that allowed state employees to buy their homes at subsidized prices, thereby affecting the wealth and the job mobility of these individuals. Based on a number of unmeasured characteristics, state employees are not equally likely to enter into self-employment. Therefore, to mitigate problems with unobserved heterogeneity Wang (2008) created a control group using a matching technique similar to the one just described. The author found that the treatment had a positive effect on entry into self-employment. Possible explanations are that becoming a house owner relaxes the credit constraint and increases job mobility.

7.2 Analysis techniques when simple conditioning is ineffective

*Use of instrumental variables***.** . Instrumental variables is one of Pearl's (2000) three basic strategies for the estimation of causal effects. An instrumental variable is actually not a variable per se but is an estimation technique that allow the researcher in a first stage to estimate a reduce form of *X'* that is then plugged in the second stage instead of the original *X'* (Wooldridge, 2002). Basically, it is a search for a variable that is exogenous to the causal system that we are interested to study but that still can function as an indicator for X. The most well-known technique is the two-stage least squares (Angrist & Krueger, 2001).

Instrumental variables are not supposed to be correlated with the sources of unwarranted heterogeniety. Instead, they are supposed to be the reduced form of the variables *X'* that we have at hand once the effect of the disturbance has been eliminated. The idea is to use another variable that is not correlated with Z but that can allow us to predict a new *X'* that is uncorrelated with Z and thereby the error term. The new *X'* has the property of not being correlated with Z (it is exogenous to the causal system), and therefore we can estimate its causal effect in our model. The main problem is to find a good instrumental variable, because the author must convince the audience that the instrument is correlated with the X of interest but not with the omitted variable Z.

There are some interesting examples of research in entrepreneurship using instrumental variables. We take two examples from the entrepreneurial finance literature[5]. Mollica and Zingales (2007) and Sorensen (2007) both address the possible effect that venture capitalists have on the ventures in which they invest. The challenges here are the problems related to omitted variables and possible reverse causality. The performance of the new venture can be caused by other factors (not measured), such the ability of the entrepreneurs, that are difficult to measure and which have a likely effect on both venture performance and on obtaining venture capital financing. It is also possible that the causality is the reverse: venture capitalists select successful firms and these firms would have been successful independent of them receiving venture financing. Although the two studies approach the problems differently, they both use instrument variables.

Mollica and Zingales (2007) use a more traditional approach, which functions as a good introductory example. They analyze whether the level of venture capital finance in a region has a positive impact on the local innovativeness (measured as numbers of patents) and new venture creation. The problem is reverse causality and that supply of innovation is difficult to control for. First, they use the citation index of published papers in a region as a proxy for supply of innovations. Second they use the size of state pension fund's assets as an instrument variable for the size of the local venture capital market. The reason is that state pension funds are often required to invest some of their money in the local venture capital funds.

Sorensen (2007) uses a somewhat more complicated approach. He is interested in explaining whether the ability of the venture capitalist increases the probability of a new venture making an IPO. He develops a sorting model where the best venture capitalists are matched to the best ventures and then applies structural estimation. As an instrument variable,

[5] We would like to thank Per Stromberg for the suggestions.

he uses the geographic structure of the VC market at a certain time. The idea is that the number of venture capital firms in one region and their quality has an effect on the sorting mechanism where certain ventures are matched to certain venture capitalist, but it should not have a direct effect on how a specific venture capitalist affects the performance of a portfolio venture. Both papers find that venture capital funding has a positive impact on IPO, new venture creation and innovativeness.

Instrument variables represent a promising avenue for research in entrepreneurship. However, authors like Morgan and Winship (2007) warn that the use of instrument variables is based on strong statistical assumptions and that the strength of an instrument is always problematic (the instrument's correlation with the original X but not with the unobserved Z).

Mediation analysis. Mediation exits when a third variable intervenes between the focal independent variable and the dependent variable. It is also one of Pearl's (2000) three basic strategies for the estimation of causal effects. The causal effect of X is transported through a system of other variable until it affects the outcome variable. Mediation analysis is the analysis and estimation of this system. Mediation analysis is relatively less frequent in entrepreneurship research while being very important in, for example, the team literature (Mathieu et al, 2008). One entrepreneurship application is Keh's et al. (2007) study of the role of Entrepreneurial Orientation on the performance of Singaporean small businesses. They found that the effect of EO on performance was in part mediated by information acquisition and information utilization. We predict that the increased interest in the entrepreneurship process and use of longitudinal data (Ireland et al., 2005) will lead to increased use of mediator effects. Mediator effects are important for examining how inputs into a process are transformed by mediators to an output. They are closely linked to the analysis of indirect and direct effects found in structural equation models (Fornell & Larcker, 1981). An obvious application is the understanding of how intraindividual processes (like the motivation system) mediates individual and contextual differences on outcomes relevant for entrepreneurship. Similarly with moderator effects, there is a debate on how to best test for mediating effects (Kenny, 2008; Wood et al., 2008).

7.3 Analysis Techniques for Dealing with Unobserved Heterogeneity

*Sensitivity analysis***.** The case may be that although no control or proxy variable can be used, the researcher wants to check the possible consequence of an unknown source of unobserved heterogeneity. Sensitivity analysis (Rosenbaum & Rubin, 1983; cf. Rosenbaum, 2002, 2004) asks the question of how much unobserved bias would need to be present to render plausible the null hypothesis of no effect. The goal is to assess how sensitive the estimates are to changes due to unobservable heterogeneity. This is somewhat different from traditional robustness checks, but this approach might be very fruitful to entrepreneurship research, since a strong and unifying theory is lacking.

The specifics of the analysis depend on the model used. Imbens (2003) and Lin et al. (1998) provide examples of techniques for sensitivity analysis developed for different models. The former develops a model with a binary treatment and a continuous dependent variable, whereas the latter develop a model with a binary treatment for hazard models and binary response data. In both cases the idea or test logic is similar. The bounds of the design is checked by testing how high the correlation can be between Z and X, and Z and Y, respectively, before it substantively changes the value of the estimated coefficient for X. If the design is robust the estimation of the X should tolerate large changes in the correlations generated by Z. Basically, the different values that Z can take are simulated in order to see how the coefficient of X changes.

There are two advantages to sensitivity analysis. First, it allows the researcher to more systematically examine how robust the reported empirical results are to changes in design.

Second, it allows better estimates of the bounds of a specific research design. A better design is less sensitive and therefore generates a more robust estimate. Sensitivity analysis allows comparison across studies that make different theoretical assumptions about the same relationship and as a consequence include different sets of control variables or apply different design solutions. Hence, widespread use of this technique would facilitate accumulation of knowledge about how sensitive previous results are to new theoretical developments, and whether a common body of knowledge across different fields in entrepreneurship can be developed.

Longitudinal Analysis Techniques. Recent reviews of entrepreneurship research show that longitudinal data collection is infrequent but increasing, so that now some 15-20 percent of published studies in high tier outlets use a longitudinal design (Dean, Shook & Payne, 2007; Coviello & Jones, 2004; Ireland et al., 2005). One advantage of using longitudinal designs is reduction of the unobserved heterogeneity problem as individuals can serve as their own controls. The causal effect can then be estimated as the change in the pre-test and the post-test measurement of their outcome (Winship and Morgan, 1999).

Longitudinal research often makes the crucial assumption that the outcome would have remained unchanged in the absence of the treatment or implied causal effect. This is often an unrealistic assumption, so we need to be able to estimate how the outcomes would have evolved in the absence of the presumed causal variable. Winship and Morgan (1999) suggest two possible sources of information to construct the counterfactual argument. First, if there are multiple pre-test observations, we can extrapolate from these observations and estimate what the outcome would have been in the absence of the causal effect. This assumes that the future is similar to the past. In any case it indicates the need for multiple measurements of outcomes before and after treatment to test sophisticated models such as the fixed effect and random effect models presented below. Second, they suggest the use of control groups. The evolution of the outcome for the control group can be used to model what the outcome would have been in the absence of the presumed causal variable. Here, the assumption is that the control group and the treatment group are similar in key respects. This is the point behind the different matching procedures that we present further down (Rubin and Waterman, 2006). This means that the researcher needs to have a deep understanding of the underlying behavior of the treatment and the control group to properly model suspected unobservable effects. So longitudinal studies might provide good choices for studying entrepreneurship, but they require long period of observation and a homogenous sample.

*Fixed and random effects methods***.** The total variance in the entrepreneurial process can be considered as the sum of the within-process variance and between-process variance. If we can assume that the unobserved factors leading to between process variance are either randomly distributed across units of observations or stable over time, we can access some powerful analytic tools known as random effect and fixed effect methods respectively. The random effects approach emphasizes between process variance while the fixed effects approach emphasizes within process variance. These methods are useful both for continuous (for example, firm performance) and dichotomous (for example, firm termination) outcome variables.

There is considerable debate about when fixed and random effect estimation, respectively, is more appropriate. Some authors, like Allison (2005), Halaby (2004) and Beck, Brüderl and Woywode (2008) argue in favor of fixed effect estimation, because it excludes the influence of the between subject variance to examine the effect of the within subject variance. Disciplinary traditions also seem to play a role. According to Halaby (2004), fixed effect estimation dominates in economics, whereas sociologists prefer random effect estimation. The same author suggests a number of techniques that allow the researcher to combine the advantages of fixed and random effects.

The following three differences between these estimation techniques might inform the choice (Balgati, 2002). The first concerns the assumption made about the unit specific error term. The fixed effect model assumes it is constant over time and specific to the unit. Put differently, the unobserved heterogeneity can be assumed to vary only across cases. The fixed effect model then just accounts for the changes across time and partials out the unexplained variance as error. A drawback is that the effect of time invariant variables such as ethnicity, sex, or education cannot be estimated. By contrast, a random effect model allows for the test of effects that vary across units but not over time, because it assumes that the individual unobservable effect is randomly distributed across units.

A second difference concerns how much change in the key variables X is observed over time. If there is little variation over time a random effects model is often preferable. A suggestion is to compare across-cases variance and across-time variance. If the across-cases variance is notably higher than the variance across time, then random effect might be a better option.

A third difference is related to the inference the researcher wishes to make. The fixed effect approach is appropriate if the focus is on a specific set of cases and the goal is to make inference about their behavior. The random effect model can be more appropriate if the sample is randomly drawn from a well-defined population. Balgati (2002) mentions the case of a panel of households. For such a sample, fixed effects estimation consumes many degrees of freedom by creating a separate intercept (i.e., akin to a dummy variable) for each case. This is an important drawback of the fixed effect model.

A technical criterion that informs the decision of which specification to use is the Hausman test. However, both Balgati (2002) and Wooldridge (2002) strongly argue that the use of that test is not sufficient for determining the choice between fixed and random effects modeling.

Interesting entrepreneurship applications using fixed effects include the work by Strahan and collaborators, which shows that a deregulated and competitive banking sector facilitates funding of start-ups (Black & Strahan, 2006; Cetorelli & Strahan, 2006). Another example is Delmar and Shane's (2003) work on the effect of business planning on new venture development. Using fixed-effect regression they are able to better isolate the effect of business planning on new venture development.

This example concludes or treatment of approaches to dealing with problematic heterogeneity at the analysis stage. The advice we have provided is summarized in Table 4.

+++++++++++++++++++++++++

Insert Table 4 about here!

+++++++++++++++++++++++++

8 CONCLUSION

In this article we have discussed three types of heterogeneity problems faced by entrepreneurship researchers: *unobserved heterogeneity* – that excluded variables may bias results – *causal heterogeneity* – that relationships may be different for different parts of the studied population – and *uneven validity* – that varying validity of operationalizations may be misinterpreted as substantive differences in relationships. When piled up on top on one another these problems may easily give the impression that conducting useful entrepreneurship research is an impossible task, especially if we are interested in estimating a causal effect. However, while perfect research is not achievable we argue that use of the strategies, 'tools' and approaches we have discussed can lead to major improvements in internal (and therefore external) validity of entrepreneurship studies. We have suggested specific remedies to heterogeneity problems in relation to causal effect estimation applicable in theory and theorizing; in research design (including sampling); in operationalization of constructs, and at the analysis stage of the research process. Attention to these remedies should allow entrepreneurship researchers to design more robust studies and to get more valid results out of existing data sets.

While following our advice should prove helpful it must be acknowledged that finding the best solution to a heterogeneity problem in a particular research situation is not just about mechanistic application of rules. According to Rosenbaum (2005), how best to deal with it is more based on creative thinking than anything else. So there is both room and need for creative, adaptive solutions in addition to application of standard recipes. Let us conclude with a few examples to illustrate this point.

Eckhardt, Shane and Delmar (2006) apply multi-stage selection modeling to the problem of predicting which new ventures will receive external funding. In the first stage they hypothesize (and estimate) that variables reflecting founders' subjective assessment of the future outlook for the venture determine whether external finance will be sought or not. In a second stage they hypothesize (and estimate) that objectively verifiable characteristics of the venture will determine external investors' willingness to fund the venture, given that financing is sought. This analytical approach is very interesting because it addresses the fundamental fact that entrepreneurship requires human agency (Shane, 2003). In cases where other variables can have their effects only if the entrepreneur so decides, this approach may be more valid than, e.g., estimating interactions between individual- and venture characteristics. Samuelsson and Davidsson's (2008) application of Latent Growth Modeling (Muthén, 1997) also deserves mention here. As this method estimates differences both in initial state and development over time it is uniquely well suited to the nascent entrepreneur sample they analyze, as one of the problems in that type of data is that the emerging ventures are unequally far progressed when they first enter the sample (cf. our above discussion of control variables). Finally, a particularly impressive example of modeling heterogeneity is Gimeno, Folta, Cooper and Woo's (1997) study of venture survival. In attacking this question they develop solutions to the problem that general human capital, as reflected in education level and management experience, should not only have a positive influence on a venture's economic performance, but also on the attractiveness of other alternatives (such as employment or starting a different venture instead). By combining elements of two known techniques (Tobit modeling and grouped data regression), each of which were novelties in entrepreneurship research, they could conclude that management experience is associated with a higher (unobserved) threshold level for acceptable performance, which explains its otherwise confusing, insignificant relationship with likelihood of exit.

Fortunately, customization and combination of approaches to heterogeneity problems do not have to reach the level of creativity and skill demonstrated by Gimeno et al. (1997) in order to be useful. For example, Chandler et al. (2008) simply but fruitfully combine moderated regression and separate, sub-sample analysis to show that predictions based on Transaction Cost Economics (Williamson, 1975) regarding the relationship between growth in employment and sales of young firms, respectively, hold up in environments that are resource scarce but not in those that are munificent. We hope our review of heterogeneity problems and their remedies can inspire entrepreneurship researchers to creatively use the techniques and follow the examples we have highlighted, thereby squeezing more valid, cumulative knowledge out of their entrepreneurship research efforts.

Figure 1

A Graphical Representation of Possible Sources of Heterogeneity Problems

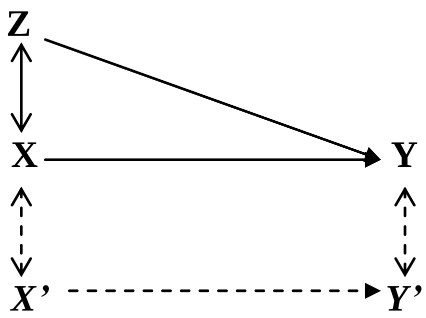

Figure 2

A Simple Schema for Theorizing about Possible Heterogeneity Issues

ΔX	I ΔY	II $\sim\Delta Y$
$\sim\Delta X$	IV ΔY	III $\sim\Delta Y$

Table 1
Dealing with heterogeneity through theory and theorizing

Establish what is already known	-	Use, e.g., Edmondson and MacManus' (2007) and/or Zahra's (2007) frameworks in order to establish the status of theory regarding the studied problem. This helps determining the extent to which theory can guide sampling, variable inclusion and modeling so as to minimize heterogeneity problems.
Systematize the theorizing	-	Use a 'counterfactual' schema in order to systematize the identification of possible heterogeneity issues.
Identify sources of heterogeneity	-	Use, e.g., Whetten's (1989) framework in order to identify sources of heterogeneity, preferably starting with *Who? When?* and *Where*? and progressing into the *How?* and *Why?* issues possibly implied by the former questions.
Specify the forms of heterogeneity effects	-	Use, e.g., Johns (2006) framework in order to systematically think through the possible, specific types of differential effects that heterogeneity can lead to.
Graph the causal system to be investigated	-	Use Pearl's (2000) graph theory to better assess which variables to measure and why. There are three basic strategies to estimate a causal effect: (a) conditioning, (b) instrumental variables and (c) mediation.

Table 2
Dealing with heterogeneity in design

Simple conditioning	
Laboratory methods	- If at all possible given the nature of the research problem, consider an experimental or simulation design, at least as an intermediary step to establishing strong internal validity before (as in a 'full cycle' approach) trying to establish external validity in a natural setting.
'Natural' experiments	- When possible, opportunistically use 'natural experiment' situations as they present unique opportunities to test relationships that may not otherwise lend themselves to experimental control.
Narrow sampling	- Consider testing theory within a narrowly defined population, sampling randomly from that population rather than from the 'entire' population for which the theoretical relationships are ultimately assumed to hold.
Stratified sampling	- Examine generalizability and boundary conditions by comparing samples from several, narrowly defined populations (i.e., stratified sampling).
Sample equivalence	- When aiming at comparison of several samples for either of the above purposes, carefully consider whether samples truly are meaningfully comparable. For this purpose method approaches used in cross-cultural research can serve as inspiration.
Control variables	- Make sure the data collection includes necessary control variables. Theory and pre-testing may be suggestive for defining the necessary set of controls (see section on dealing with heterogeneity in analysis).
Unobserved heterogeneity	
Longitudinal design	- Consider a longitudinal design with multiple waves of collection of data on both independent and dependent variables in order to be able to control for case-level, idiosyncratic effects (see section on dealing with heterogeneity in analysis). Here the case functions as its own control as we measure it before and after the treatment.

Table 3
Dealing with heterogeneity in operationalization

Quality of customized measures	-	Sampling from a narrow population has the added advantage of facilitating the use of customized, high quality operationalizations.
Uneven validity I	-	When a broader sample or multiple strata are used, carefully consider and pre-test the conceptual and scaling equivalence of operationalizations. For this purpose method approaches used in cross-cultural research can serve as inspiration (including semantic equivalence if the entrepreneurship research itself is cross-cultural).
Uneven validity II	-	If a universally equally valid operationalization cannot be achieved, theoretical replication may be more effectively performed using customized operationalizations for each sub-sample.
Uneven validity III – the dependent variable(s)	-	A single DV measure is unlikely to achieve conceptual equivalence for abstracted outcome constructs (e.g., 'success'; 'performance'). Inclusion of a range of outcome measures is advised in order to gain a better understanding of the contingencies involved.

Table 4
Dealing with heterogeneity in analysis

Simple conditioning	
Moderation	These are good approaches to model and estimate causal heterogeneity as suggested by theory. However, following the example of published entrepreneurship research does not guarantee correct application or interpretation.
Post-matching procedures and propensity scoring	- Modern post-matching procedures borrow from the experimental perspective. By creating matches of cases in treatment and not in treatment, the unobserved heterogeneity problem of self-selection is taken into account
Simple conditioning ineffective	
Instrumental variables	This technique allows us to create a new *X'* variable where the effect of the disturbing Z variable has been eliminated. Hence the new *X'* is no longer correlated to the Z variable and estimations are unbiased
Mediation	There exist a third observed variable or more that are hypothesized to mediate the effect of one input variable on the dependent variable. If they can be indentified and included in the analysis we can isolate the causal system.
Unobserved heterogeneity	
Sensitivity analysis	- When unobserved heterogeneity is suspected but cannot be controlled for, sensitivity analysis helps assess how sensitive estimates are to possible sources of unobserved heterogeneity.
Fixed and random effects models	- Longitudinal data offer strong safeguards against unobserved heterogeneity because cases can serve as their own controls. The choice of fixed or random effects estimation should not be based simply on disciplinary tradition or a 'Hausman test' but be informed by the assumptions made about the unit-specific error term; the amount of variation across time vs. across cases, and what type of inference the researcher is aiming for.

References

Acs, Z. J., & Audretsch, D. B. (Eds.). (2003). *Handbook of Entrepreneurship Research: An Interdisciplinary Survey and Introduction*. Dordrecht, NL: Kluwer.

Aguinis, H. (2002). Estimation of interaction effects in organizational studies. *Organizational Research Methods, 5*(3), 207-211.

Aguinis, H. (2004) *Regression Analysis for Categorical Moderators*. New York: The Guilford Press.

Allison, P. D. 2005. *Fixed Effects Regression Methods for Longitudinal Data using SAS*. Cary, NC: SAS Institute Inc.

Alvarez, S., & Busenitz, L. (2001). The entrepreneurship of resource-based theory. *Journal of Management, 27*(6), 755.

Astley, W., & Van de Ven, A. (1983). Central perspectives and debates in organization theory. *Administrative Science Quarterly, 28*(2), 245-273.

Bacharach, S. (1989). Organizational theories: Some criteria for evaluation. *Academy of Management Review*, 496-515.

Balgati, B. H. 2001. *Econometric Analysis of Panel Data* (Second ed.). Chichester, England: John Wiley & Sons Ltd.

Baron, R. (2006). Opportunity recognition as pattern recognition: How entrepreneurs "connect the dots" to identify new business opportunities. *Academy of Management Perspectives, 20*(1), 104-119.

Baron, R., & Ward, T. (2004). Expanding entrepreneurial cognition's toolbox: Potential contributions from the field of cognitive science. *Entrepreneurship Theory and Practice, 28*(6), 553-573.

Baron, R. M., & Kenny, D. A. (1986). The mediator-moderator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology, 89*(4), 587-598.

Beck, N., Brüderl, J. & Woywode, M. (2008). Momentum or deceleration? Theoretical and methodological reflections on the analysis of organizational change. *Academy of Management Journal*, 51(3): 413-435.

Black, S. E., & Strahan, P. E. (2006). Entrepreneurship and bank credit availability. *Journal of Finance, 57*(6), 2807-2833.

Bouckenooghe, D., De Clercq, D., Willem, A., & Buelens, M. (2007). An assessment of validity in entrepreneurship research. *Journal of Entrepreneurship, 16*(2), 147-171.

Brambor, T., Clark, T. W., & Golder, M. (2006). Understanding interaction models: Improving empirical analysis. *Political Analysis, 14*(1), 63-82.

Brown, T., Davidsson, P., & Wiklund, J. (2001). An operationalization of Stevenson's conceptualization of entrepreneurship as opportunity-based firm behavior. *Strategic Management Journal, 22*(10), 953-968.

Brundin, E., Patzelt, H., & Shepherd, D. (2008). Managers'emotional displays and employees' willingness to act entrepreneurially. *Journal of Business Venturing, 23*(2), 221-243.

Byrne, B., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155.

Cetorelli, N., & Strahan, P. E. (2006). Finance as barriers to entry: Bank competition and industry structure in local U.S. markets. *Journal of Finance, 61*(1), 437-461.

Chandler, G. N., & Lyon, D. W. (2001). Issues of research design and construct measurement in entrepreneurship research. *Entrepreneurship Theory & Practice, 25*(4, Summer), 101-113.

Chandler, G. N., McKelvie, A., & Davidsson, P. (2008). Asset specificity and behavioral uncertainty as moderators of the sales growth — employment growth relationship in emerging ventures. *Journal of Business Venturing*, available online 21 May 2008.
Chatman, J. A., & Flynn, F. J. (2005). Full-cycle micro-organizational behavior research. *Organization Science, 16*, 434-447.
Cialdini, R. B. (1980). Full cycle social psychology. In L. Beckman (Ed.), *Applied Social Psychology Annual* (Vol. 1, pp. 21-48). Beverly Hills, CA.: Sage.
Cliff, J. E., Devereaux-Jennings, P., & Greenwood, R. (2006). New to the game and questioning the rules: The experiences and beliefs of founders who start imitative versus innovative firms. *Journal of Business Venturing, 21*, 633-663.
Coviello, N. E., & Jones, M. V. (2004). Methodological issues in international entrepreneurship research. *Journal of Business Venturing, 19*, 485-508.
Davidsson, P. (2004). *Researching Entrepreneurship*. New York: Springer.
Davidsson, P. (2006). Nascent entrepreneurship: Empirical studies and developments. *Foundations and Trends in Entrepreneurship, 2*(1), 1-76.
Davidsson, P., & Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing, 18*(3), 301-331.
Davis, J., Eisenhardt, K., & Bingham, C. (2007). Developing theory through simulation methods. *The Academy of Management Review, 32*(2), 480-499.
Dean, M. A., Shook, C. L., & Payne, T. G. (2007). The past, the present and future of entrepreneurship research: Data analystic trends and training. *Entrepreneurship Theory and Practice, 2007*(July), 601-618.
Delmar, F., & Shane, S. (2006). Does experience matter? The effect of founding team experience on the survival and sales of newly founded ventures. *Strategic Organization, 4*(3), 215-247.
Dickson, K., & Kangaraarachchi, R. (2006). Sri Lankan experiences of post-tsunami small business survival and recovery. Paper presented at *British Academy of Management Annual Conference*, Belfast, September 2006.
Edmondson, A. S., & McManus, S. (2007). Methodological fit in management field research. *Academy of Management Review, 32*(4), 1155-1179.
Eckhardt, J., Shane, S., & Delmar, F. (2006). Multistage selection and the financing of new ventures. *Management Science, 52*(2), 220-232.
Eisenhardt, K. M., & Schoonhoven, C. B. (1990). Organizational growth: Linking founding team, strategy, environment and growth among US semiconductor ventures, 1978-1988. *Administrative Science Quarterly, 35*(3), 504-530.
Fearon, J. D. (1991). Counterfactuals and hypothesis testing in political science. *World Politics*, 43(2, Jan.): 169-195.
Fiet, J., Piskounov, A., & Patel, P. (2005). Still Searching (Systematically) for Entrepreneurial Discoveries. *Small Business Economics, 25*(5), 489-504.
Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*, 39-50.
Galbraith, C., & Stiles, C. (2006). Disasters and entrepreneurship: A short review. In C. Galbraith & C. Stiles (Eds.), *Developmental Entrepreneurship: Adversity, Risk, and Isolation*. New York: Elsevier.
Gartner, W. B., Carter, N. M., & Reynolds, P. D. (2004). Business start-up activities. In W. B. Gartner, K. G. Shaver, N. M. Carter & P. D. Reynolds (Eds.), *Handbook of Entrepreneurial Dynamics: The Process of Business Creation* (pp. 285-298). Thousand Oakes: Sage.
Gartner, W. B., Shaver, K. G., Carter, N. M., & Reynolds, P. D. (2004). *Handbook of Entrepreneurial Dynamics: The Process of Business Creation*. Thousand Oaks, CA: Sage.

Gimeno, J., Folta, T. B., Cooper, A. C., & Woo, C. Y. 1997. Survival of the fittest? Entrepreneurial human capital and the persistence of underperforming firms. *Administrative Science Quarterly*, 42: 750-783.

Gregoire, D., Barr, P. S., & Shepherd, D. (forthcoming). Cognitive processes of opportunity recognition: The role of structural alignment. *Organization Science*

Gustafsson, V. (2006). *Entrepreneurial Decision-Making*: *Tasks and Cognitions*. Cheltenham, UK and Northampton, MA.: Elgar.

Halaby, C. N. 2004. Panel models in sociological research: Theory into practice. *Annual Review of Sociology*, 30: 507-544.

Hamilton, R. (1989). Unemployment and business formation rates: Reconciling time series and cross section evidence. *Environment and Planning, 21*, 249-255.

Hammond, K. R. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics, 17*(5).

Harrison, R. T., Mason, C. M., & Girling, P. (2004). Financial bootstrapping and venture development in the software industry. *Entrepreneurship and Regional Development, 16*, 307-333

Hausman, J., & Taylor, W. E. 1981. Panel data and unobservable individual effects. *Econometrica*, 49(6): 1377-1398.

Heckman, J. J. 1979. Sample selection bias as a specification error. *Econometrica*, 47(1): 153-162.

Heckman, J. J. 2000. Causal parameters and policy analysis in economics: A twentieth century retrospective. *The Quarterly Journal of Economics*, 115(1): 45-97.

Heckman, J.J., & Navarro-Lozano, S. 2004. Using matching, instrumental variables, and control functions to estimate economic choice models. *The Review of Economics and Statistics*, 86(1): 30.

Henrekson, M. (2007). Entrepreneurship and institutions. *Comparative Labor Law & Policy Journal*, *28*(3).

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal, 28*, 331-343.

Höfler, M. (2005). Causal inference based on counterfactuals. *BMC Medical Research Methodology*, 5(28): 1-12.

Hubbard, R., Vetter, D. E., & Little, E. L. (1998). Replication in strategic management: Scientific testing for validity, generalizability, and usefulness. *Strategic Management Journal, 19*, 243-254.

Imbens, G. W. 2003. Sensitivity to exogeneity assumptions in program evaluation. *The American Economic Review*, 93(2): 126-132.

Ireland, D., & Webb, J. W. (2007). A Cross-disciplinary exploration of entrepreneurship research. *Journal of Management, 33*(6), 891.

Ireland, D., Webb, J. W., & Coombs, J. e. (2005). Theory and methodology in entrepreneurship research. In D. J. Ketchen & D. D. Bergh (Eds.), *Research Methodology in Strategy and Management* (Vol. 2). New York: Elsevier.

Jaccard, J., & Turrisi, R. (2003). *Interaction Effects in Logistic Regression* (2 ed.): Sage Publications Inc.

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review, 31*(2), 386-408.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica, 50*(3), 649-670.

Keh, H. T., Nguyen, T. T. M., & Ng, H. P. 2007. The effects of entrepreneurial orientation and marketing information on the performance of SMEs. *Journal of Business Venturing*, 22(4): 592-611.

Kenny, D. A. 2008. Reflections on mediation. *Organizational Research Methods*, 11(2): 353.

Ketchen, D. J., Boyd, B. K., & Bergh, D. D. (2008). Research methodology in strategic management: Past accomplishments and future challenges. *Organizational Research Methods, 11*, 643-658.

Kim, P. H., Aldrich, H. E., & Keister, L. A. (2006). Access (not) denied: The impact of financial, human, and cultural capital on entrepreneurial entry in the United States. *Small Business Economics, 27*(1), 5-22.

King, G., Keohane, R., & Verba, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

Kish, L. (1987). *Statistical Design for Research*. New York: Wiley.

Kleinknecht, A., Van Montfort, K., & Brouwer, E. (2002). The non-trivial choice between innovation indicators. *Economics of Innovation & New Technology, 11*(2), 109-122.

Kreiser, P., Marino, L., & Weaver, K. (2002). Assessing the Psychometric Properties of the Entrepreneurial Orientation Scale: A Multi-Country Analysis. *Entrepreneurship: Theory and Practice, 26*(4), 71-95.

Koppl, R., & Minniti, M. (2003). Market processes and entrepreneurial studies. In Z.J. Acs & D. B. Audretsch (Eds.), *Handbook of Entrepreneurship Research*. Dordrecht: Kluwer.

Lawless, M., & Tegarden, L. (1991). A Test of performance similarity among strategic group members in conforming and non-conforming industry structures. *Journal of Management Studies, 28*(6), 645-664.

Lebow, R. N. (2000). What's so different about a counterfactual? *World Politics*, 52: 550-585.

Lee, L. (1982). Specification error in multinomial logit models: Analysis of the omitted variable bias. *Journal of Econometrics, 20*(2).

Lewis, D. (1973). Counterfactuals and comparative possibility. *Journal of Philosophical Logic, 2*(4), 418-446.

Lin, D. Y., Psaty, B. M., & Kronmal, R. A. 1998. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. *Biometrics*, 54(3): 948-963.

Lindh, T., & Ohlsson, H. 1996. Self-employment and windfall gains: Evidence from the Swedish lottery. *The Economic Journal*, 106(439): 1515-1526.

Low, M. B., & MacMillan, I. C. (1988). Entrepreneurship: past research and future challenges. *Journal of Management, 14*(2), 139-161.

Lumpkin, G. T., & Dess, G. G. (1996). Clarifying the entrepreneurial orientation construct and linking it to performance. *Academy of Management Review, 21*(1), 135-172.

Maillat, D. (1998). Innovative milieux and new generations of regional policies. *Entrepreneurship & Regional Development, 10*(1), 1-16.

March, J. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71-87.

Mathieu, J., Travis Maynard, M., Rapp, T. & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, 34(3): 410-476.

Mugler, J. (1990). Entrepreneurship and the theory of the firm. In R. Donckels & A. Miettinen (Eds.), *New findings and perspectives in entrepreneurship* (pp. 3-15). Aldershot: Avebury.

Murphy, G. G., Trailer, J. W., & Hill, R. C. (1996). Measuring performance in entrepreneurship research. *Journal of Business Research, 36*, 15-23.

Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data. In A. Raftery (Ed.), *Socoilogical Methodology*. Boston, MA: Blackwell Publishers.

Nopo, H. & Valenzuela, P. (2007). Becoming an Entrepreneur. *IZA Discussion Paper* No. 2716 Available at SSRN: http://ssrn.com/abstract=981702

Nelson, R. R., & Winter, S. G. (1982). *An Evolutionary Theory of Economic Change*. Cambridge, MA.: Belknap Press.

Oakes, M. (1986). *Statistical Inference: A Commentary for the Social and Behavioural Sciences.* Chichester: Wiley.
Organizational Research Methods (2002), Vol. 5; Iss. 3 (special issue on moderated regression).
Pearl, J. (2000). *Causality: Models, Reasoning and Inference.* NY, USA: Cambridge University Press.
Popper, K. (ed.) (1959). *The Logic of Scientific Discovery*. New York, NY: Basic Books.
Rauch, A., Wiklund, J., Lumpkin, T, & Frese, M. (2006). *Entrepreneurial orientation and business performance: A meta-analysis.* Working Paper. Giessen: Justus-Liebig University.
Reynolds, P. D. (2007). New firm creation in the US: A PSED overview. *Foundations and Trends in Entrepreneurship, 3*(1), 1-151.
Reynolds, P. D., & Curtin, R. T. (2008). Business creation in the United States: Panel Study of Entrepreneurial Dynamics II initial assessment. *Foundations and Trends in Entrepreneurship, 4*(3), 155-307.
Rosenbaum, P. R. 2002. *Observational Studies*. New York: Springer.
Rosenbaum, P. R. 2004. Design sensitivity in observational studies. *Biometrika*, 91(1): 153.
Rosenbaum, P. R. 2005. Heterogeneity and causality: Unit heterogeneity and design sensitivity in observational studies. *The American Statistician*, 59(2): 147-152.
Rosenbaum, P. R., & Rubin, D. B. 1983. The central role of the propensity score in observational studies of causal effects. *Biometrika*, 70: 41-55.
Rubin, D. B., & Waterman, R. P. (2006). Estimating the causal effects of marketing interventions using propensity score methodology. *Statistical Science, 21*(2), 206-222.
Runyan, R. C. (2006). Small business in the face of crisis: Identifying barriers to recovery from a natural disaster. *Journal of Contingencies and Crisis Management, 14*(1), 12-26.
Samuelsson, M., & Davidsson, P. (2008). Does venture opportunity variation matter? Investigating systematic process differences between innovative and imitative new ventures. *Small Business Economics, (pre-published electronically by Springer at www.springerlink.com.ezp01.library.qut.edu.au/content/100338*)
Sarasvathy, S. (2001). Causation and effectuation: towards a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review, 26*(2), 243-288.
Sarasvathy, S. (2008). *Effectuation: Elements of Entrepreneurial Expertise*. Cheltenham, UK and Northampton, MA: Elgar.
Schade, C. (2005). Dynamics, experimental economics, and entrepreneurship. *Journal of Technology Transfer*, 30: 409-431.
Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods, 6*, 169-215.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies*, 597-619.
Shane, S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science,* 11(4): 448-469.
Shane, S. 2003. *A General Theory of Entrepreneurship. The Individual-Opportunity Nexus*. Northampton, MA: Edward Elgar.
Shane, S. (2006). Introduction to the focused issue on entrepreneurship. *Management Science, 52*(2), 155-159.
Shepherd, D., & DeTienne, D. (2005). Prior knowledge, potential financial reward, and opportunity identification. *Entrepreneurship Theory and Practice, 29*(1), 91-112.
Shepherd, D., & Zacharakis, A. (1997). Conjoint Analysis: A window of opportunity for entrepreneurship research. In J. Katz & R. H. Brockhaus (Eds.), *Advances in*

Entrepreneurship, Firm Emergence, and Growth (Vol. 3, pp. 203-248). Greenwich, CT: JAI Press.
Shugan, S. M. 2006. Errors in the variables, unobserved heterogeneity, and other ways of hiding statistical error. *Marketing Science*, 25(3): 203-216.
Sorensen, O. & Stuart, T.E. (2008). Entrepreneurship: A field of dreams?. *The Academy of Management Annals*, 2 (1): 517-543.
Stevenson, H. H., & Jarillo, J. C. (1990). A paradigm of entrepreneurship: Entrepreneurial management. *Strategic Management Journal, 11*, 17-27.
Stuart, T. E., Huang, H., & Hybels, R. 1999. Interorganizational endorsements and the performance of entrepreneurial ventures. *Administrative Science Quarterly*, 44: 315-349.
Taylor, M. P. 2001. Self-employment and windfall gains in Britain: Evidence from panel data. *Economica*, 68(272): 539-565.
Usher, J. M., & Evans, M. G. (1996). Life and death along gasoline alley: Darwinian and Lamarckian processes in a differentiating population. *Academy of Management Journal, 39*(5), 1428-1466.
Venkataraman, S. (1997). The distinctive domain of entrepreneurship research: An editor's perspective. In J. Katz & J. Brockhaus (Eds.), *Advances in Entrepreneurship, Firm Emergence, and Growth* (Vol. 3, pp. 119-138). Greenwich, CT: JAI Press.
Walter, A., Auer, M., & Ritter, T. (2006). The impact of network capabilities and entrepreneurial orientation on university spin-off performance. *Journal of Business Venturing, 21*, 541-567.
Wang, S.-Y. 2008. Credit Constraints, Job Mobility and Entrepreneurship: Evidence From a Property Reform in China. Berkeley University, CIG Working Paper No. 84. Accessed at *http://igov.berkeley.edu/workingpapers/papers0708.html*
Weick, K. (1989). Theory construction as disciplined imagination. *Academy of Management Review, 14*(4), 516-531.
Wennekers, S., Stel, A., Thurik, A. R., & Reynolds, P. D. (2005). Nascent entrepreneurship and the level of economic development. *Small Business Economics, 24*, 293-309.
Western, B. (1998). Causal heterogeneity in comparative research: A Bayesian hierarchical modeling approach. *American Journal of Political Science, 42*, 1233-1259.
Whetten, D. A. (1989). What constitutes a theoretical contribution? *Academy of Management Review, 14*(4), 490-495.
Williamson, O. (1975). *Markets and Hierarchies*. New York: The Free Press.
Winship, C. & Morgan, S.L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, 25:659-707.
Wood, R. E., Goodman, J. S., Beckmann, N., & Cook, A. 2008. Mediation testing in management research: A review and proposals. *Organizational Research Methods*, 11(2): 270.
Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: The MIT Press.
Yang, D. D. (2006). Empirical social inquiry and models of causal inference. *The New England Journal of Political Science, 2*(1), 51-88.
Zahra, S. A. (2007). Contextualizing theory building in entrepreneurship research. *Journal of Business Venturing, 22*(3), 443-452.
Zahra, S. A., & Hayton, J. (2008). The effect of international venturing on firm performance: The moderating influence of absorptive capacity. *Journal of Business Venturing, 23*, 195-220.